



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genetic Risk Factors of ME/CFS

Citation for published version:

Dibble, JJ, McGrath, SJ & Ponting, CP 2020, 'Genetic Risk Factors of ME/CFS: A Critical Review', *Human Molecular Genetics*. <https://doi.org/10.1093/hmg/ddaa169>

Digital Object Identifier (DOI):

[10.1093/hmg/ddaa169](https://doi.org/10.1093/hmg/ddaa169)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Human Molecular Genetics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Genetic Risk Factors of ME/CFS: A Critical Review

Journal:	<i>Human Molecular Genetics</i>
Manuscript ID	HMG-2020-R2-00484.R1
Manuscript Type:	4 Invited Review Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Dibble, Joshua; University of Edinburgh Western General Hospital, MRC Human Genetics Unit Ponting, Chris; University of Oxford, MRC Human Genetics Unit
Key Words:	Myalgic Encephalomyelitis, Genome-wide association studies, Complex disease, Chronic Fatigue Syndrome

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Genetic Risk Factors of ME/CFS: A Critical Review

Joshua J. Dibble¹, Simon J. McGrath², and Chris P. Ponting^{1,*}

¹ MRC Human Genetics Unit,
Institute of Genetics and Molecular Medicine,
University of Edinburgh,
Crewe Road South, Edinburgh, EH4 2XU, United Kingdom
² Wrekin, Prospect Road, Monmouth, NP25 3SZ, United Kingdom

* Corresponding author:
Tel: +44 (0) 131 651 8500
email: chris.ponting@igmm.ed.ac.uk

For Peer Review

Abstract

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a complex multisystem illness that lacks effective therapy and a biomedical understanding of its causes. Despite a prevalence of approximately 0.2–0.4% and its high public health burden, and evidence that it has a heritable component, ME/CFS has not yet benefited from the advances in technology and analytical tools that have improved our understanding of many other complex diseases. Here we critically review existing evidence that genetic factors alter ME/CFS risk before concluding that most ME/CFS candidate gene associations are not replicated by the larger CFS cohort within UK Biobank. Multiple genome-wide association studies of this cohort also have not yielded consistently significant associations. Ahead of upcoming larger genome-wide association studies we discuss how these could generate new lines of enquiry into the DNA variants, genes and cell-types that are causally involved in ME/CFS disease.

Introduction

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a long-term, multisystem illness of unknown aetiology whose symptoms are highly debilitating (1). Large numbers of individuals are affected, estimated at 0.2% of the UK population (2), with the ratio of women to men diagnosed as high as 4-to-1 (2,3). This results in a substantial annual economic burden assessed at approximately £3.3 billion annually in the UK (<https://meassociation.org.uk/wp-content/uploads/2020Health-Counting-the-Cost-Sept-2017.pdf>). Those diagnosed with ME/CFS experience diverse physical and cognitive symptoms but are most distinguished on the basis of post-exertional malaise, defined as a substantial worsening of symptoms following mental or physical activity that to healthy individuals would be minor, and which are not alleviated by resting (1, 4). The quality of life of people with ME/CFS is worse than for many other illnesses, including some cancers (5).

Individuals with ME/CFS can be diagnosed after failing to recover fully following acute infection from viral and non-viral pathogens (6). In such cases it is assumed that infection initiates ME/CFS and thus contributes to its aetiology. Overall, however, the biological basis to ME/CFS is poorly understood, and currently there is little consensus among investigators on the molecular, cellular and genetic factors that alter risk of ME/CFS (7). Many biomolecular studies have sought biomarkers (factors such as RNAs, microRNAs, metabolites or proteins) whose abundance in blood, for example, accurately and reproducibly distinguishes ME/CFS cases from controls (8-14). Unfortunately, such a factor has yet to be identified in multiple independent studies. If it were to be found then it would also need to accurately separate ME/CFS cases from people with unrelated diseases.

Multiple members of the same family can be diagnosed with ME/CFS (15, 16) which implies that inherited (i.e. genetic) factors might contribute to ME/CFS risk. However, ME/CFS diagnoses do not follow a predictable Mendelian pattern, which reveals that there is not one genetic variant that increases ME/CFS risk. Instead, ME/CFS is likely to be a complex multifactorial disorder whose genetic contributions are many and varied, as is known to be the case for example for many autoimmune diseases (17). These varied contributions will likely be of small individual effect, but together disrupt many genes and cellular or physiological processes. Each such variant is expected to contribute little to altered risk, implying that even in aggregate they cannot provide a reliable diagnosis. Genetics can objectively identify genes, molecules and cellular pathways that contribute to ME/CFS genetic risk, and these can then be targeted therapeutically. By contrast, molecular differences observed between patients and controls in other biomolecular studies could reflect the many secondary consequences of disease (e.g. a more sedentary lifestyle) rather than the disease's primary cause.

Here we first summarise ME/CFS diagnosis criteria before detailing the evidence for ME/CFS heritability and genetic risk factors. We then review the methodology, limitations and advantages of genome-wide association studies as applied to ME/CFS. This review is timely because it precedes the launch of DecodeME (www.decodeme.org.uk) and a similar study from the University of Oslo (www.ukbiobank.ac.uk/tag/cfs-me/).

ME/CFS case definitions

Population genetics studies of disease rely on accurate case definition. For ME/CFS, however, no laboratory diagnostic test is currently approved. Rather, clinical diagnosis is made based on physical examination, case history and exclusion of other disorders. For research studies, investigators do not currently apply a single set of diagnostic criteria for ME/CFS or CFS cases.

Of the 20 existing sets of diagnostic criteria or case definitions (1), three are commonly applied in research and the use of a fourth (IOM/NAM) is expected in the future (**Box 1**). These different diagnostic criteria will select distinct, albeit overlapping, case cohorts and thus may not yield equivalent biomolecular findings.

Evidence that CFS risk is inherited

Various observations are consistent with genetic factors contributing to CFS risk for some individuals. Individuals with a CFS diagnosis (Fukuda or ICD-9 code 780.71 criteria) have a significant excess relatedness over the wider population for both close (first or second degree) and distant (third degree) relatives (16, 22). Of three studies that have estimated narrow-sense heritability (h^2) using large cohorts, two reported non-zero h^2 -values that provide evidence for heritability of risk for CFS and, presumably, ME/CFS. An analysis of USA health insurance claimed a high narrow-sense heritability ($h^2 = 0.48$) of CFS (23) whereas an analysis of UK Biobank individuals self-reporting a CFS diagnosis reported a less striking heritability (single nucleotide polymorphism- [SNP-] based approximate $h^2 = 0.08$ with low confidence) (24, <http://www.nealelab.is/uk-biobank>). The third, a large twin-based study of CFS-like cases, produced an inconclusive result, with the 95% confidence interval of h^2 including zero [0.03 (0.00–0.65)] (25).

Mitochondrial and HLA genetics

Independent studies confirm that clinically proven mtDNA variants do not commonly explain ME/CFS (26-28). People with ME/CFS, however, appear more likely to carry mtDNA that lacks even mildly deleterious variants (28), a finding whose implications require further investigation.

HLA proteins enable the immune system to differentiate self- from non-self cells such as foreign pathogens. Their genes exhibit extreme population polymorphism and certain HLA types are genetically predisposed to particular autoimmune diseases (29). Two independent HLA types tagged by HLA-C*07:04 or HLA-DQB1*03:03 were recently shown to be significantly associated with ME/CFS status (Canadian consensus criteria) (30). These alleles are each carried by ~10% of ME/CFS individuals and alter risk by approximately 1.5–2.0-fold. If these results are independently replicated then they indicate that genetic differences in the human immune system alter risk for ME/CFS.

Genome-Wide Association Study (GWAS)

A GWAS is ideal for discovering genetic causes of disease and new biology particularly when disease aetiology is unknown, as is the case for ME/CFS. This is not just because it is comprehensive but because its results are not influenced by pre-existing biological assumptions or hypotheses. GWAS was central to unexpected findings, for example of the role of glia, rather than neurons, in Alzheimer's disease pathogenesis (31) and of components of the interleukin-23 pathway in ankylosing spondylitis, psoriasis and psoriatic arthritis (32). Despite GWAS being expensive, sales of medications that have benefitted from this method already exceed its costs (32), and these are declining rapidly.

In a GWAS approximately 0.5–2.5 million SNPs are genotyped and a stringent threshold of the test statistic (p -value) applied: $< 5 \times 10^{-8}$ for SNPs with minor allele frequency [MAF] $> 5\%$, or $< 1 \times 10^{-8}$ for MAF $\geq 0.1\%$ (33, 34). For CNVs the p -value threshold is less stringent because they are fewer in number, although no broad consensus on its value has yet been reached. To be adequately powered to find DNA variants of small effect, GWAS requires SNP genotype data from large numbers of cases and controls (preferably from many more controls than cases).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Case sample sizes of approximately 10^2 are well powered only to identify common (MAF $\sim 10\%$) variants with very strong effects (effect size ~ 1) whereas sample sizes of 10^4 are required to identify MAF $\sim 10\%$ variants with weaker effects (~ 0.1) (32). This explains why GWAS often employ cohort sizes of $\sim 10^4$ - 10^6 , and why when studies increase their cohort sizes they tend to discover larger numbers of lower-effect loci (35).

ME/CFS genetics and UK Biobank

UK Biobank is a 500,000-strong cohort that has been genotyped and exceptionally well phenotyped and hence is well suited to GWAS (36). One of the over 750 anthropometric and disease-related traits captured by UK Biobank is self-reported clinical diagnosis of CFS. This generated the largest cohort of people collected to date who have self-reported a clinical diagnosis (1,829 people). This number was approximately 0.45% of all UK Biobank participants and showed a marked female-bias (0.61% of females and 0.26% of males), as expected (37). As there is a likely ‘healthy volunteer’ selection bias (38) in UK Biobank these prevalence estimates are lower-bound values. It is unknown how many of these 1,829 participants meet the ME/CFS criteria above, hence all results discussed below are preliminary and, as ever, require statistically robust replication.

Five groups have performed a case-control GWAS on CFS cases in UK Biobank. Unfortunately, they reach no consensus and their results are far from being definitive. One study found no DNA variant to pass the $p < 5 \times 10^{-8}$ threshold (39). Another study reported results both partitioned and unpartitioned by sex (<http://www.nealelab.is/uk-biobank>) finding one variant associated with female CFS status and another with male or female CFS status (**Table 1**). Two other studies (24, 40) highlighted a single variant each (**Table 1**). The final study (<https://pan.ukbb.broadinstitute.org/>) found no variants associated with either the

Biobank CFS phenotype (**Table 1**), or the “phecode” definition of CFS (41). Consequently, despite these five studies analysing the same data set, not a single associated DNA variant was replicated by multiple analyses. Two explanations of this lack of replication are likely. First, as alleles become rarer the likelihood increases that all people with the minor allele are placed among the cases just by chance. Two of the highlighted SNPs that are very rare in the population ($MAF < 0.5\%$) are possible examples of this. Second, DNA sites that appear to have three or more alleles are usually excluded from analyses because they likely reflect technical genotyping artefacts (<http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UK-Biobank-Axiom-Array-Content-Summary-2014-1>). One of the highlighted variants (rs148723539) is one such multi-allelic site (**Table 1**).

The single remaining variant (rs7337312) is neither rare ($MAF \approx 0.5$) nor multi-allelic and was identified in the female-only CFS GWAS by Neale *et al.* (<http://www.nealelab.is/uk-biobank>). This variant, together with adjacent significant variants that are in linkage disequilibrium, occur within a 51kb region containing the *SLC25A15* gene (**Figure 1**). In most cases the gene affected by DNA variation is not the nearest (42) because of long-range genetic regulation (43). Nonetheless, because the rs7337312 variant predicts the amount of *SLC25A15* mRNA transcribed from this gene in some tissue samples, *SLC25A15* could be a causal gene of altered CFS risk. *SLC25A15* encodes the Ornithine Transporter type 1 (ORNT1) protein that transports ornithine (as well as lysine and arginine) across the inner membrane of mitochondria to the mitochondrial matrix. Ornithine is an amino acid that plays a role in the urea cycle. A person with the rs7337312 CFS risk allele is expected to produce lower amounts of *SLC25A15* mRNA resulting in reduced transport of ornithine into the mitochondrion and higher amounts of ornithine in blood. Yamano *et al.* (44) and Naviaux *et al.* (14), but not Armstrong *et al.* (45), report some evidence in support of this prediction.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Using the same UK Biobank data, Aguirre *et al.* (46) tested for association with CNVs, finding two genes (*TCOF1* and *THUMPD2*) to have a greater number of CNVs (either gains or losses) in UK Biobank CFS cases than in controls after applying a multiple-testing significance threshold of $p < 3.1 \times 10^{-6}$. Nevertheless, these results need to be treated with caution owing to CNV calling artefacts being prevalent at the low allele frequencies ($<0.1\%$) used in the study.

Studies not using UK Biobank data

Smith *et al.* (47) undertook a GWAS on CFS (defined using Fukuda criteria) for very low numbers (40 cases and 40 non-fatigued control subjects) and reported 65 DNA variants as being associated at a non-standard highly permissive threshold of $p < 10^{-3}$. When the well-established threshold of $p < 5 \times 10^{-8}$ is applied, however, none remain significant. This is expected because, as discussed above, GWAS using this number of cases are only well powered to identify population frequent alleles with strong effects.

A second GWAS also used similarly small cohort sizes (42 cases, defined using both Canadian and Fukuda criteria, and 38 controls) and reported 299 DNA variants as associated with ME/CFS status at $p < 1 \times 10^{-5}$ (48). The authors justified this permissive threshold as being an inclusion criterion of the GWAS Catalog (49). Nevertheless, $p < 1 \times 10^{-5}$ is used by the GWAS Catalog for reporting purposes only, and only for results from the overall (initial GWAS plus replication) population, when this study lacked a replication cohort. 15 variants were reported as genome-wide significant at $p < 5 \times 10^{-8}$ but these associations are not replicated in the UK Biobank cohort (**Figure 2**).

Perez et al. (50) conducted a genetics study using 383 people who self-reported a clinical diagnosis of ME/CFS (Fukuda criteria). The study only considered DNA variants that disrupted genes despite >90% of variants associated with diseases or traits lying elsewhere in the genome (51). Using non-standard thresholds on variant frequency, the authors reported 5,693 DNA variants that are two-fold more (or less) frequent in cases versus controls. However, such case-control allele frequency inequalities could have two technical, rather than biological, explanations. Firstly, errors stemming from poor DNA sample quality, incomplete DNA hybridisation to the genotyping array, or poorly performing array probes. Controlling for these errors imperfectly, or inconsistently between case and control genotypes, has led to retraction of publications reporting genetic associations (for example, (52)). Secondly, errors arising because confounding effects from the controls such as differences in genetic ancestry were not accounted for.

Candidate gene studies

A recent review by leading human geneticists stated that: “initial efforts targeting variants within ‘candidate’ genes were plagued by inadequate power, unduly liberal thresholds for declaring significance and scant attention to sources of bias and confounding, resulting in overblown claims and failed replication” (53). Their confidence stems from knowledge that tens of thousands of associations have been identified by GWAS that subsequently are often replicated independently (49). An example of a candidate gene approach in ME/CFS research is a study of 9 DNA variants in the gene *NR3C1* for 40 people with CFS and 42 controls (54). Four of these variants (rs1866388, rs2918419, rs860458 and rs6188) passed their statistical threshold for significance ($p < 0.05$). Nevertheless, none of these variants survive a Bonferroni multiple testing correction ($p < 0.05/9$ or $p < 0.006$) and none are replicated with the UK Biobank CFS cohort (24) ($p = 0.96, 0.71, 0.71$ and 0.97 , respectively).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

To generalise this point, we again exploited the large cohort of self-reported CFS participants of UK Biobank. We tested for replication a large set of genetic findings reported in 16 CFS studies published between 2003 and 2015 that met the systematic review criteria of Wang *et al.* (55). Looking up associations between DNA variants (Table 1 of reference (55)) and CFS status in UK Biobank (24) for the 11 studies that had readily available SNP references yielded a p -value distribution between 0 and 1 (**Figure 2a**). Replication would require these p -values to be skewed towards small values. No such skew is observed (**Figure 2a**), consistent with random samples, hence these initial findings show no evidence of being replicated.

This conclusion is further substantiated from plotting the replication p -values for a further set of 77 DNA variants from Marshall-Gradisnik *et al.* (56), and 23 from Schlauch, *et al.* (48): these p -values also show a lack of skew towards small values (**Figure 2b**).

Expected outcomes of a ME/CFS GWAS

GWAS are proposed to have “substantially improved our understanding of the mechanisms responsible for many rare and common diseases and driven development of novel preventative and therapeutic strategies” (53). This suggests that large GWAS on ME/CFS are overdue. Replicated results from such studies would have four important implications.

Firstly, it would catalyse the gain of much-needed insight into genes, cellular processes and tissues or cell types that causally alter risk for ME/CFS. When combined with functional genomics and other technologies (53), a well-designed GWAS can pinpoint multiple chromosomal locations containing DNA variants that change the activity of genes – in specific cells or tissues – that thereby alter a person’s risk of ME/CFS. If these genes are known to have

an activity in common – such as a mitochondrial or neurological or immunological function – then this common feature prioritises cellular processes and molecular mechanisms that could be causally involved in disease. Framing such causal hypotheses has been aided considerably by the knowledgebase of gene function, including activity levels, molecular mechanism and cellular function, which have been growing substantially and rapidly over recent years as a result of novel and higher throughput technologies.

Secondly, a GWAS would enable detection of genetic signals that ME/CFS shares with other diseases or traits. Methods (e.g. (57)) that compare GWAS summary statistics for ME/CFS and other traits are available to calculate the genetic correlations between them. Genetic signals for ME/CFS could be shared with other diseases just as autoimmune diseases (for example, rheumatoid arthritis, type 1 diabetes, and autoimmune thyroid disease) share such signals and underlying mechanisms of disease (58).

Thirdly, a GWAS could aid stratification of ME/CFS sub-types. Despite their well-defined clinical diagnoses, complex diseases such as type 2 diabetes are caused by diverse molecular and cellular mechanisms (59) and this should also be expected of ME/CFS. Its underlying biological sub-types could eventually be detectable using methods that test for genetic effect heterogeneity (60).

Lastly, discovery of genetic factors for ME/CFS risk might be expected to improve how this disorder is perceived by health professionals and by society at large.

Future perspective

Genetics studies are the best way to understand the aetiology of ME/CFS, due to the causal nature of genetic associations. A large GWAS focused on discovering the biomolecular mechanisms of ME/CFS is urgently needed because no study on the genetics of ME/CFS yet has seen results repeated under replication. For an appropriately powered GWAS, at least 10⁴ participants are required, and an equal or greater number of controls. A strict $p < 5 \times 10^{-8}$ or 1×10^{-8} statistical significance threshold must also be applied to reduce the numerous false positive associations seen from the meta-analyses presented here.

Although recruiting thousands of people with ME/CFS – particularly severely affected individuals who are house- or bed-bound – is a challenging task, it will be essential to perform a GWAS using their samples if we are to understand the mechanisms of the disease. With case criteria refined using genetic findings it may then be possible to begin stratifying the disease into distinct sub-types each with a different causal mechanism and potentially a specific treatment.

Acknowledgments

This work was supported by the Medical Research Council [MC_UU_00007/15 to C.P.P.]; Action for ME and the Chief Scientist Office, Scotland [AME/CSO/18/01]. We also thank the Medical Research Council and National Institute for Health Research for funding the DecodeME GWAS project [MC_PC_20005]. People with ME/CFS can now register to participate in this project at <https://www.decode-me.org.uk/>.

References

1. Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (2015) Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness. National Academies Press (US), Washington DC.
2. Nacul,L.C., Lacerda,E.M., Pheby,D., Champion,P., Molokhia,M., Fayyaz,S., Leite,J.C.D.C., Poland,F., Howe,A. and Drachler,M.L. (2011) Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: A repeated cross-sectional study in primary care. *BMC Med.*, **9**, 91.
3. Valdez,A.R., Hancock,E.E., Adebayo,S., Kiernicki,D.J., Proskauer,D., Attewell,J.R., Bateman,L., DeMaria,A., Lapp,C.W., Rowe,P.C. et al. (2019) Estimating prevalence,

demographics, and costs of ME/CFS using large scale medical claims data and machine learning. *Front. Pediatr.*, **6**, 412.

4. Jason,L.A., Kot,B., Sunnquist,M., Brown,A., Reed,J., Furst,J., Newton,J.L., Strand,E.B. and Vernon,S.D. (2015) Comparing and contrasting consensus versus empirical domains. *Fatigue*, **3**, 63-74.

5. Hvidberg,M.F., Brinth,L.S., Olesen,A.V., Petersen,K.D. and Ehlers,L. (2015) The health-related quality of life for patients with myalgic encephalomyelitis / chronic fatigue syndrome (ME/CFS). *PLoS One*, **10**, e0132421.

6. Hickie,I., Davenport,T., Wakefield,D., Vollmer-Conna,U., Cameron,B., Vernon,S.D., Reeves,W.C. and Lloyd,A. (2006) Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: Prospective cohort study. *Br. Med. J.*, **333**, 575.

7. Edwards, J.C.W., McGrath, S., Baldwin, A., Livingstone, M. and Kewley, A. (2016), The biological challenge of myalgic encephalomyelitis/chronic fatigue syndrome: a solvable problem. *Fatigue*, **4**, 63-69.

8. Missailidis,D., Annesley,S.J., Allan,C.Y., Sanislav,O., Lidbury,B.A., Lewis,D.P. and Fisher,P.R. (2020) An isolated complex V inefficiency and dysregulated mitochondrial function in immortalized lymphocytes from ME/CFS patients. *Int. J. Mol. Sci.*, **21**, 1074.

9. Gow,J.W., Hagan,S., Herzyk,P., Cannon,C., Behan,P.O. and Chaudhuri,A. (2009) A gene signature for post-infectious chronic fatigue syndrome. *BMC Med. Genomics*, **2**, 38.

10. Kerr,J.R., Petty,R., Burke,B., Gough,J., Fear,D., Sinclair,L.I., Matthey,D.L., Richards,S.C.M., Montgomery,J., Baldwin,D.A. et al. (2008) Gene Expression Subtypes in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *J. Infect. Dis.*, **197**, 1171-1184.

11. VanElzakker,M.B., Brumfield,S.A. and Lara Mejia,P.S. (2019) Neuroinflammation and cytokines in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): A critical review of research methods. *Front. Neurol.*, **9**, 1033.

12. Almenar-Pérez,E., Sarriá,L., Nathanson,L. and Oltra,E. (2020) Assessing diagnostic value of microRNAs from peripheral blood mononuclear cells and extracellular vesicles in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Sci. Rep.*, **10**, 2064.

13. Fluge,Ø., Mella,O., Bruland,O., Risa,K., Dyrstad,S.E., Alme,K., Rekeland,I.G., Sapkota,D., Røslund,G.V., Fosså,A. et al. (2016) Metabolic profiling indicates impaired pyruvate dehydrogenase function in myalgic encephalopathy/chronic fatigue syndrome. *JCI Insight*, **1**, e89376.

14. Naviaux,R.K., Naviaux,J.C., Li,K., Bright,A.T., Alaynick,W.A., Wang,L., Baxter,A., Nathan,N., Anderson,W. and Gordon,E. (2016) Metabolic features of chronic fatigue syndrome. *Proc. Natl. Acad. Sci. U. S. A.*, **113**, E5472-E5480.

15. Walsh,C.M., Zainal,N.Z., Middleton,S.J. and Paykel,E.S. (2001) A Family History Study of Chronic Fatigue Syndrome. *Psychiatr. Genet.*, **11**, 123-128.

16. Underhill,R.A. and O'Gorman,R. (2011) Prevalence of chronic fatigue syndrome and chronic fatigue within families of CFS patients. *J CFS*, **13**, 3-13.
17. Li,Y.R., Li,J., Zhao,S.D., Bradfield,J.P., Mentch,F.D., Maggadottir,S.M., Hou,C., Abrams,D.J., Chang,D., Gao,F. et al. (2015) Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat. Med.*, **21**, 1018-1027.
18. Fukuda,K., Straus,S.E., Hickie,I., Sharpe,M.C., Dobbins,J.G. and Komaroff,A. (1994) The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. *Ann. Intern. Med.*, **121**, 953-959.
19. Carruthers,B.M., Jain,A.K., De Meirleir,K.L., Peterson,D.L., Klimas,N.G., Lemer,A.M., Bested,A.C., Flor-Henry,P., Joshi,P., Powles,A.C.P. et al. (2003) Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome: Clinical working case definition, diagnostic and treatment protocols. *J CFS*, **11**, 7-115.
20. Jason,L.A., Brown,A., Clyne,E., Bartgis,L., Evans,M. and Brown,M. (2012) Contrasting Case Definitions for Chronic Fatigue Syndrome, Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Myalgic Encephalomyelitis. *Eval. Health Prof.*, **35**, 280-304.
21. Carruthers,B.M., Van de Sande,M.I., De Meirleir,K.L., Klimas,N.G., Broderick,G., Mitchell,T., Staines,D., Powles,A.C.P., Speight,N., Vallings,R. et al. (2011) Myalgic encephalomyelitis: International Consensus Criteria. *J. Intern. Med.*, **270**, 327-338.
22. Albright,F., Light,K., Light,A., Bateman,L. and Cannon-Albright,L.A. (2011) Evidence for a heritable predisposition to chronic fatigue syndrome. *BMC Neurol.*, **11**, 62.
23. Lakhani,C.M., Tierney,B.T., Manrai,A.K., Yang,J., Visscher,P.M. and Patel,C.J. (2019) Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes. *Nat. Genet.*, **51**, 327-334.
24. Canela-Xandri, O., Rawlik, K. and Tenesa, A. (2018) An atlas of genetic associations in UK Biobank. *Nat. Genet.*, **50**, 1593-1599.
25. Sullivan, P.F., Evengård, B., Jacks, A. and Pedersen, N.L. (2005) Twin analyses of chronic fatigue in a Swedish national sample. *Psychol. Med.*, **35**, 1327-1336.
26. Schoeman,E.M., Van Der Westhuizen,F.H., Erasmus,E., van Dyk,E., Knowles,C.V.Y., Al-Ali,S., Ng,W.F., Taylor,R.W., Newton,J.L. and Elson,J.L. (2017) Clinically proven mtDNA mutations are not common in those with chronic fatigue syndrome. *BMC Med. Genet.*, **18**, 29.
27. Billing-Ross,P., Germain,A., Ye,K., Keinan,A., Gu,Z. and Hanson,M.R. (2016) Mitochondrial DNA variants correlate with symptoms in myalgic encephalomyelitis/chronic fatigue syndrome. *J. Transl. Med.*, **14**, 19.
28. Venter,M., Tomas,C., Pienaar,I.S., Strassheim,V., Erasmus,E., Ng,W.F., Howell,N., Newton,J.L., Van der Westhuizen,F.H. and Elson,J.L. (2019) MtDNA population variation in Myalgic encephalomyelitis/Chronic fatigue syndrome in two populations: a study of mildly deleterious variants. *Sci. Rep.*, **9**, 2914.

29. Matzaraki, V., Kumar, V., Wijmenga, C. and Zhernakova, A. (2017) The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.*, **18**, 76.
30. Lande, A., Fluge, Ø., Strand, E.B., Flåm, S.T., Sosa, D.D., Mella, O., Egeland, T., Saugstad, O.D., Lie, B.A. and Viken, M.K. (2020) Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Sci. Rep.*, **10**, 5267.
31. Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y. et al. (2019) Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*, **366**, 1134-1139.
32. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. and Yang, J. (2017), 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.*, **101**, 5-22.
33. Fadista, J., Manning, A.K., Florez, J.C. and Groop, L. (2016) The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.*, **24**, 1202-1205.
34. Wu, Y., Zheng, Z., Visscher, P.M. and Yang, J. (2017) Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.*, **18**, 86.
35. López-Cortegano, E. and Caballero, A. (2019) Inferring the nature of missing heritability in human traits using data from the GWAS catalog. *Genetics*, **212**, 891-904.
36. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, **12**, e1001779.
37. Bakken, I.J., Tveito, K., Gunnes, N., Ghaderi, S., Stoltenberg, C., Trogstad, L., Håberg, S.E. and Magnus, P. (2014) Two age peaks in the incidence of chronic fatigue syndrome/myalgic encephalomyelitis: A population-based registry study from Norway 2008-2012. *BMC Med.*, **12**, 167.
38. Taylor, A.E., Jones, H.J., Sallis, H., Euesden, J., Stergiakouli, E., Davies, N.M., Zammit, S., Lawlor, D.A., Munafò, M.R., Smith, G.D. et al. (2018) Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.*, **47**, 1207-1216.
39. Tanigawa, Y., Li, J., Justesen, J.M., Horn, H., Aguirre, M., DeBoever, C., Chang, C., Narasimhan, B., Lage, K., Hastie, T. et al. (2019) Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight adipocyte biology. *Nat Commun.*, **10**, 4064.
40. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A. et al. (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.*, **50**, 1335-1341.

41. Wei,W.Q., Bastarache,L.A., Carroll,R.J., Marlo,J.E., Osterman,T.J., Gamazon,E.R., Cox,N.J., Roden,D.M. and Denny,J.C. (2017) Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One*, **12**, e0175508.
42. Gusev,A., Ko,A., Shi,H., Bhatia,G., Chung,W., Penninx,B.W.J.H., Jansen,R., De Geus,E.J.C., Boomsma,D.I., Wright,F.A. et al. (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, **48**, 245-252.
43. Kleinjan,D.A. and Van Heyningen,V. (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.*, **76**, 8-32.
44. Yamano,E., Sugimoto,M., Hirayama,A., Kume,S., Yamato,M., Jin,G., Tajima,S., Goda,N., Iwai,K., Fukuda,S. et al. (2016) Index markers of chronic fatigue syndrome with dysfunction of TCA and urea cycles. *Sci. Rep.*, **6**, 34990.
45. Armstrong,C.W., McGregor,N.R., Sheedy,J.R., Buttfield,I., Butt,H.L. and Gooley,P.R. (2012) NMR metabolic profiling of serum identifies amino acid disturbances in chronic fatigue syndrome. *Clin. Chim. Acta*, **413**, 1525-1531.
46. Aguirre,M., Rivas,M.A. and Priest,J. (2019) Phenome-wide Burden of Copy-Number Variation in the UK Biobank. *Am. J. Hum. Genet.*, **105**, 373-383.
47. Smith,A.K., Fang,H., Whistler,T., Unger,E.R. and Rajeevan,M.S. (2011) Convergent genomic studies identify association of GRIK2 and NPAS2 with chronic fatigue syndrome. *Neuropsychobiology*, **64**, 183-194.
48. Schlauch,K.A., Khaiboullina,S.F., De Meirleir,K.L., Rawat,S., Petereit,J., Rizvanov,A.A., Blatt,N., Mijatovic,T., Kulick,D., Palotás,A. et al. (2016) Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Transl. Psychiatry*, **6**, e730.
49. Buniello,A., Macarthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. et al. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005-D1012.
50. Perez,M., Jaundoo,R., Hilton,K., Alamo,A.D., Gemayel,K., Klimas,N.G., Craddock,T.J.A. and Nathanson,L. (2019) Genetic predisposition for immune system, hormone, and metabolic dysfunction in myalgic encephalomyelitis/chronic fatigue syndrome: A pilot study. *Front. Pediatr.*, **7**, 206.
51. Maurano,M.T., Humbert,R., Rynes,E., Thurman,R.E., Haugen,E., Wang,H., Reynolds,A.P., Sandstrom,R., Qu,H., Brody,J. et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.
52. Sebastiani,P., Solovieff,N., Puca,A., Hartley,S.W., Melista,E., Andersen,S., Dworkis,D.A., Wilk,J.B., Myers,R.H., Steinberg,M.H. et al. (2010) Genetic Signatures of Exceptional Longevity in Humans. *Science*, **333**, 404.

53. Claussnitzer,M., Cho,J.H., Collins,R., Cox,N.J., Dermitzakis,E.T., Hurles,M.E., Kathiresan,S., Kenny,E.E., Lindgren,C.M., MacArthur,D.G. et al. (2020) A brief history of human disease genetics. *Nature*, **577**, 179-189.
54. Goertzel,B.N., Pennachin,C., de Souza Coelho,L., Gurbaxani,B., Maloney,E.M. and Jones,J.F. (2006) Combinations of single nucleotide polymorphisms in neuroendocrine effector and receptor genes predict chronic fatigue syndrome. *Pharmacogenomics*, **7**, 475-483.
55. Wang,T., Yin,J., Miller,A.H. and Xiao,C. (2017) A systematic review of the association between fatigue and genetic polymorphisms. *Brain Behav. Immun.*, **62**, 230-244.
56. Marshall-Gradisnik,S., Johnston,S., Chacko,A., Nguyen,T., Smith,P. and Staines,D. (2016) Single nucleotide polymorphisms and genotypes of transient receptor potential ion channel and acetylcholine receptor genes from isolated B lymphocytes in myalgic encephalomyelitis/chronic fatigue syndrome patients. *J. Int. Med. Res.*, **44**, 1381-1394.
57. Bulik-Sullivan,B., Finucane,H.K., Anttila,V., Gusev,A., Day,F.R., Loh,P.R., Duncan,L., Perry,J.R.B., Patterson,N., Robinson,E.B. et al. (2015) An atlas of genetic correlations across human diseases and traits. *Nat. Genet.*, **47**, 1236-1241.
58. Richard-Miceli,C. and Criswell,L.A. (2012) Emerging patterns of genetic overlap across autoimmune disorders. *Genome Med.*, **4**, 6.
59. Scott,R.A., Scott,L.J., Mägi,R., Marullo,L., Gaulton,K.J., Kaakinen,M., Pervjakova,N., Pers,T.H., Johnson,A.D., Eicher,J.D. et al. (2017) An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*, **66**, 2888-2902.
60. Dahl,A., Cai,N., Ko,A., Laakso,M., Pajukanta,P., Flint,J. and Zaitlen,N. (2019) Reverse GWAS: Using genetics to identify and model phenotypic subtypes. *PLoS Genet.*, **15**, e1008009.

Legends to Figures

Figure 1.

A ME/CFS-associated GWAS locus on chromosome 13 (X-axis). Genome-wide significant SNPs from Neale et al. <http://www.nealelab.is/uk-biobank> are those lying above the horizontal dashed line. The location of the *SLC25A15* gene is indicated below (exons are shown as vertical bars). The left-hand Y-axis reflects the statistical confidence ($-\log_{10}$ p-value) of the association between DNA variant and ME/CFS case status. The right-hand Y-axis (blue data curves) indicates the low extent of recombination within this locus. rs7337312 is highlighted as the reference variant (purple diamond). The degree of linkage disequilibrium of rs7337312 with neighbouring SNPs is indicated as r^2 (right-hand legend).

Figure 2.

Histograms of replication p -values for association of SNPs for CFS risk. These p -values were obtained from a UK Biobank CFS GWAS (24). Only variants tested in the UK Biobank were considered.

Panel a. Association p -values for SNPs identified in reference (59).

Panel b. Association p -values for SNPs identified in references (60) and (52).

Box 1. ME/CFS Case Inclusion/Exclusion Criteria.

- 1) The Fukuda Criteria developed in 1994 by the International Chronic Fatigue Syndrome Study Group (18). To be diagnosed with CFS, a person must display unexplained, persistent or relapsing chronic fatigue, as well as at least four of eight additional symptoms.
- 2) The Canadian Consensus Criteria (19). To be diagnosed, people with ME/CFS must exhibit a broad array of symptoms in specific combinations, namely:
 - i. cardinal symptoms of fatigue, post-exertional malaise or post-exertional fatigue, sleep dysfunction and pain (myalgia, often including headaches);
 - ii. two or more neurological and/or cognitive symptoms from a given list;
 - iii. at least one symptom from two of the following categories: autonomic, neuroendocrine and immune manifestations (with a list of relevant symptoms provided for each); and,
 - iv. symptoms must have persisted for at least six months.The authors also provided a list of co-morbidities, and another of illnesses that exclude the diagnosis of ME/CFS. These criteria seek to define patients with diverse symptoms, and have been proposed to preferentially select individuals with more severe symptoms (20).
- 3) International Consensus Criteria (21): This modifies the Canadian Consensus Criteria by revising criteria, adding others and removing both chronic fatigue as a criterion and the requirement of a six-month period prior to diagnosis. For diagnosis, a patient must display:
 - i. post-exertional neuroimmune exhaustion (broadly equivalent to post-exertional malaise);
 - ii. three neurological impairment symptoms at least one each from three of four categories;
 - iii. three immune, gastro-intestinal or genitourinary symptoms at least one from three of five categories; and,
 - iv. at least one energy production/transportation symptom.
- 4) USA Institute of Medicine (IOM, now called the National Academy of Medicine [NAM]) Criteria, developed for clinical not research purposes (1). Criteria were streamlined and focussed on core symptoms:
 - i. substantial reduction from pre-illness activity levels, evidence of post-exertional malaise and unrefreshing sleep; and,
 - ii. cognitive impairment and/or orthostatic intolerance.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

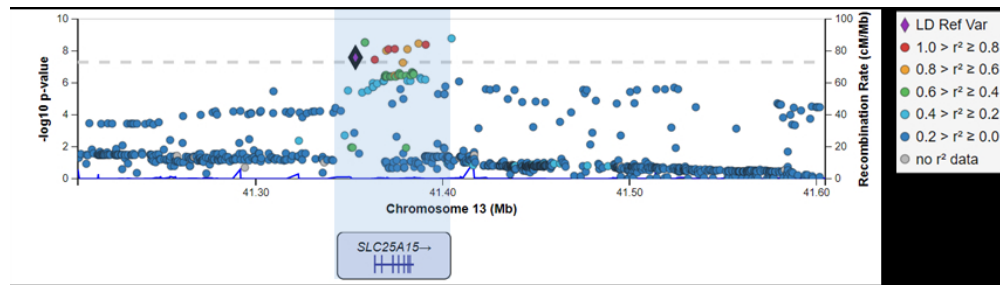
Table

Table 1 - Summary of SNPs identified as significant in the UK Biobank CFS Cohort. Bold font indicates nominally significant *p*-values (< 5x10⁻⁸). These five studies are cited in the Main Text.

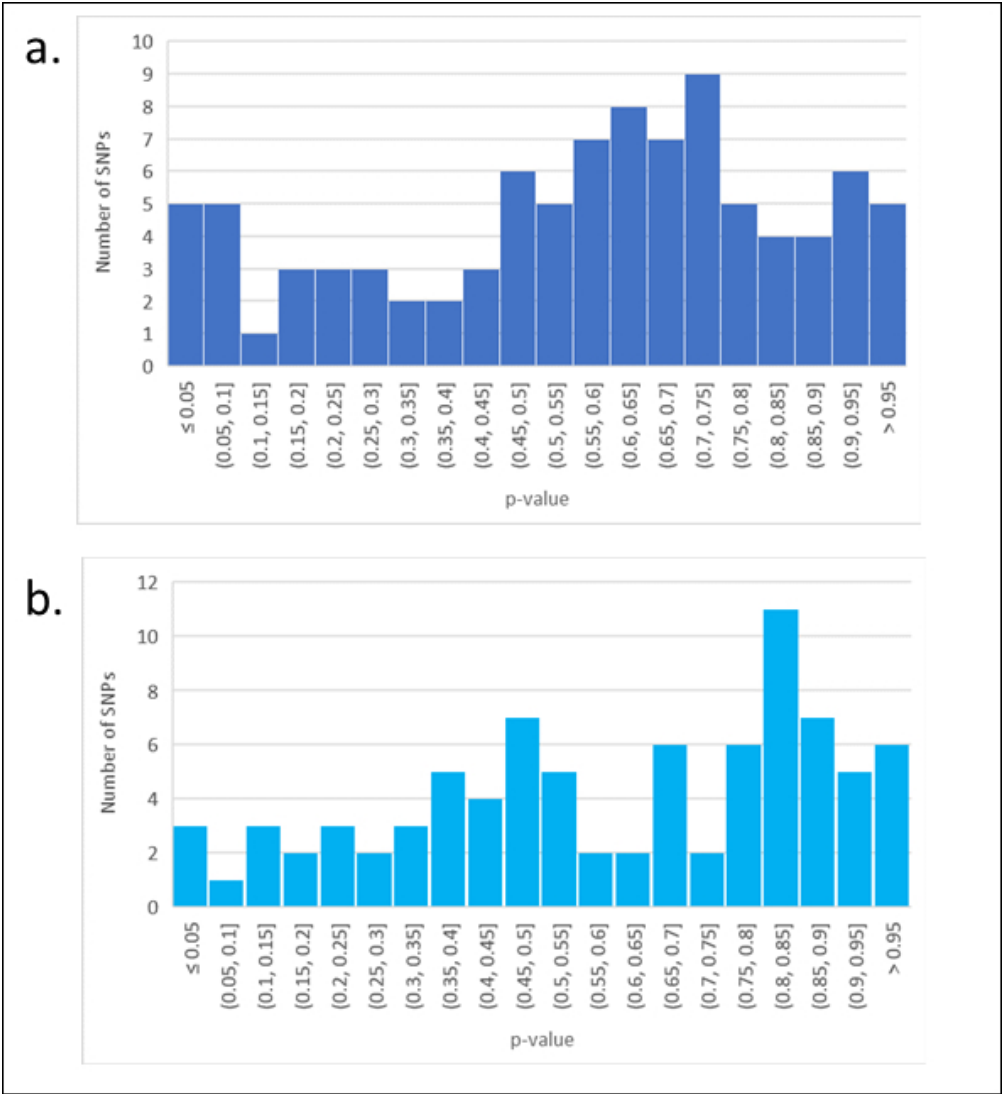
DNA Variant	Chromosome	Nearby Gene	Minor Allele Freq (GnomAD)	<i>p</i> (Neale) female	<i>p</i> (Neale) male	<i>p</i> (Neale) both	<i>p</i> (SAIGE) both	<i>p</i> (GeneAtlas) both	<i>p</i> (Global Biobank Engine)	<i>p</i> (Pan-UKBB)
rs7337312	13	SLC25A15	0.54	2.6x10⁻⁸	0.74	4.0x10 ⁻⁶	0.29	0.00017	9.3x10 ⁻⁶	2.1x10 ⁻⁴
rs150954845	10	P4HA1	0.00029	not reported	not reported	not reported	0.37	2.6x10⁻¹²	not reported	not reported
rs148723539	10	EBF3	0.0093	9.7x10 ⁻⁷	6.1x10 ⁻⁴	2.3x10⁻⁹	0.29	not reported	not reported	3.7x10 ⁻⁶
rs564809936	4	COX7B2	0.00016	not reported	not reported	not reported	3.8x10⁻⁸	not reported	not reported	4.8x10 ⁻³

Abbreviations

- CNV – Copy number variant
- GWAS – Genome wide association study
- HLA – Human Leukocyte Antigen
- IOM/NAM – USA Institute of Medicine (now called the National Academy of Medicine [NAM])
- MAF – minor allele frequency
- ME/CFS – Myalgic Encephalomyelitis/Chronic Fatigue Syndrome
- mt – mitochondrial
- ORNT1 – Ornithine Transporter type 1
- SNP – single nucleotide polymorphism



A ME/CFS-associated GWAS locus on chromosome 13 (X-axis). Genome-wide significant SNPs from Neale et al. <http://www.nealelab.is/uk-biobank> are those lying above the horizontal dashed line. The location of the SLC25A15 gene is indicated below (exons are shown as vertical bars). The left-hand Y-axis reflects the statistical confidence ($-\log_{10}$ p-value) of the association between DNA variant and ME/CFS case status. The right-hand Y-axis (blue data curves) indicates the low extent of recombination within this locus. rs7337312 is highlighted as the reference variant (purple diamond). The degree of linkage disequilibrium of rs7337312 with neighbouring SNPs is indicated as r^2 (right-hand legend).



Histograms of replication p-values for association of SNPs for CFS risk. These p-values were obtained from a UK Biobank CFS GWAS (24). Only variants tested in the UK Biobank were considered.

Panel a. Association p-values for SNPs identified in reference (59).

Panel b. Association p-values for SNPs identified in references (60) and (52).